

# Computed optical spectra of SARS-CoV-2 proteins\*

Zhuo Li<sup>a</sup> and Jonathan D. Hirst<sup>b\*</sup>

<sup>a</sup>*School of Pharmaceutical Sciences, Jilin University, Changchun 130021, China.*

<sup>b</sup>*School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom.*

Email: [jonathan.hirst@nottingham.ac.uk](mailto:jonathan.hirst@nottingham.ac.uk)

\*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Abstract.** Treatment for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes Covid-19, may well be predicated on knowledge of the structures of protein of this virus. However, often these cannot be determined easily or quickly. Herein, we provide calculated circular dichroism (CD) spectra in the far- and near-UV, and infra-red (IR) spectra in the amide I region for experimental structures and computational models of SARS-CoV-2 proteins. The near-UV CD spectra offer greatest sensitivity in assessing the accuracy of models.

## Introduction

Since the outbreak of SARS-CoV-2 at the beginning of 2020, scientists have been seeking insights that will underpin solutions to this new pandemic. Structural characterization of the proteins of this virus is crucial in terms of understanding their biological functions, finding inhibitors and designing vaccines. The structure of SARS-CoV-2 spike glycoprotein has been determined by cryo-electron microscopy and is being used in the design of vaccines targeting entry into cells.<sup>1</sup> The main protease, which is essential in processing RNA translation, is another key target. The crystal structures of the main protease with and without inhibitors have been reported,<sup>2</sup> which have provided a basis for some theoretical studies. Wang tried to repurpose approved drugs or drugs in clinical trials by performing docking screening against the crystal structure of the main

protease.<sup>3</sup> Huynh et al. screened the ZINC database targeting the main protease as well,<sup>4</sup> but they used the structure of the protein receptor after equilibration in molecular dynamics (MD) simulations. Docking with the equilibrated structure revealed an interesting binding mechanism, the occupation of the “anchor” site by the ligand. The virus attaches to the host cell via recognition of the angiotensin-converting enzyme 2 (ACE2). Han and Král designed peptide inhibitors by mimicking the helical structure on the binding interface of ACE2 to block the binding of the SARS-CoV-2 receptor binding domain (RBD).<sup>5</sup> These inhibitors showed stable binding in MD simulations. When combining the experimental structures with MD simulations, we can get better understanding of some biological processes of SARS-CoV 2, such as the recognition of the RBD and ACE2<sup>6</sup> and the catalytic mechanism of the 3C-like protease.<sup>7</sup>

Currently, structures of more than half of the SARS-CoV-2 proteins have not yet been determined experimentally. Prediction of protein structures using computational methods is an alternative approach. Computational researchers from around the world are sharing the models of SARS-CoV-2 proteins that they obtained with different prediction algorithms, either from a homology template or de novo, in order to facilitate drug design or functional studies.<sup>8</sup> However, the predicted protein structures need to be validated via experimental methods. Optical spectroscopy may provide valuable information in this respect.

## **Methods**

In this study, we have collected SARS-CoV-2 protein coordinates either from experimental determination as reported in the Protein Data Bank (PDB) or from computational modelling<sup>9-11</sup> in order to calculate their optical spectra (Table S1 in the Supplementary Information). Eight of the proteins have more than one PDB entry. Multiple PDB structures of the same protein were used if they were determined by the same research group under similar experimental conditions. Other PDB entries of the considered proteins were not included, because of the variation of the experimental

conditions, bound ligands and complex composition. For the spike glycoprotein we only calculated its near-UV CD spectrum, due to its large size. Protein structures are flexible under physiological conditions. Although informative, a crystal structure or a modelled structure can only represent a fixed stable state. Thus, we also calculated near-UV CD spectra using snapshots from MD trajectories shared by D. E. Shaw Research<sup>12</sup> of five proteins to take into account the conformational flexibility of the proteins in solution. Simulations of the spike protein RBD-ACE2 complex (6M17), spike glycoprotein (6VXX and 6VYB) and RNA polymerase-nsp7-nsp8 (6M71) had been performed for 10  $\mu$ s. We extracted 425 snapshots sampled uniformly every 24 ns to use in the calculation of the near-UV CD spectra. The simulation of the main protease (6Y84) was over a period of 100  $\mu$ s and for this protein we extracted 500 snapshots sampled uniformly every 200 ns.

The methodology for the MD simulations was described in the report of D. E. Shaw.<sup>12</sup> We summarise just the most salient details; additional information and the associated literature references can be found in the original report.<sup>12</sup> The trajectory of the apo main protease started from the X-ray crystal structure (PDB entry 6Y84). The protein was solvated in a 120 x 120 x 120 Å water box. The simulation was conducted at 300 K in the NPT ensemble and was performed on Anton 2 using the DES-Amber force field. Trajectories of the trimeric SARS-CoV-2 spike glycoprotein were initiated from the closed state (PDB entry 6VXX), and from a partially opened state (PDB entry 6VYB). The simulations used the Amber ff99SB-ILDN force field for proteins, the TIP3P model for water, and the generalized Amber force field for glycosylated asparagine. The simulations were conducted at 310 K in the NPT ensemble. The protocol used for the spike glycoprotein was also used for the spike protein RBD-ACE2 complex. For all simulations, the systems were neutralized and salted with NaCl, with a final concentration of 0.15 M.

We considered 23 computational models of 17 protein or protein fragments from three sources (Table S1). The SWISS-MODEL server has provided the full SARS-CoV-2 proteome including 25 protein models and five hetero-oligomeric complexes.<sup>9</sup> In this

study, we selected the models where there are no experimental coordinates available and where their QMEAN quality estimates<sup>13</sup> indicate high quality/confidence. Thus, we computed spectra of seven SwissModel SARS-CoV-2 structures (Table S1). There are proteins where no homology model is available, due to a lack of a suitable template. Six proteins have also been studied with the 'free modelling' method, AlphaFold, which is a deep learning system to predict the structure of proteins for which there is no similar template on which to base a model.<sup>10</sup> Models of these proteins and another four were refined by Heo and Feig<sup>11</sup> via their pipeline including inter-residue distance prediction with trRosetta,<sup>14</sup> lowest energy model selection and MD simulation-based refinement. The secondary structure content of the proteins shown in Table S1 was calculated with the DSSP server.<sup>15,16</sup>

We calculate three types of optical spectra of SARS-CoV-2 proteins: far-ultraviolet (UV) electronic circular dichroism (CD), near-UV CD and infrared (IR) spectra in the amide I region. Although these techniques do not provide atomic level structural information, they reflect the conformational flexibility of proteins in solution (and by extension under physiological conditions). All three spectroscopic methods can monitor conformational changes perhaps induced by changes in experimental conditions or by ligand/inhibitor binding. Thus, comparing the experimental spectra with the calculated spectra may provide information about the quality of the structural model.

Far-UV CD spectroscopy is sensitive to chiral features of the protein backbone which are reflected by the excitation bands arising from the peptide bond. Each type of protein secondary structure has its unique far-UV CD features<sup>17</sup> and the far-UV CD spectrum can be deconvolved as a combination of each secondary structure component.<sup>18</sup> It has been used in determining the secondary structures of the papain-like protease of MERS<sup>19</sup> and the envelope protein of SARS.<sup>20</sup> The far-UV CD has also been used to measure the interaction between the SARS-CoV-2 spike protein receptor binding domain (RBD) and heparin.<sup>21</sup> Calcium binding by the SARS 3a protein has been

monitored using CD.<sup>22</sup> The far-UV CD spectrum showed the increase of  $\alpha$ -helicity when mixing peptides from two heptad repeat regions of the SARS spike protein which suggested the formation of a complex.<sup>23,24</sup> This fusogenic mechanism is also evident in SARS-CoV-2 and has been used to design fusion inhibitors.<sup>25</sup> In this study, we employed the PDB2CD server,<sup>26</sup> an empirical method, to predict the far-UV CD spectra of the SARS-CoV-2 proteins. A reference dataset, SMP180 (soluble + membrane), is used to search structural similarity of the query protein and proteins in the dataset and the CD spectra of the reference proteins are used to create the CD spectrum.

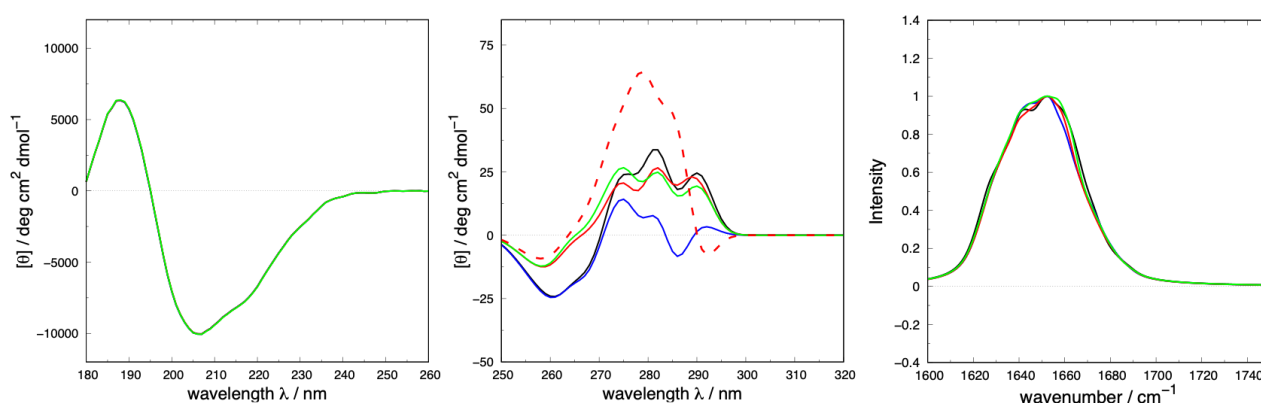
The amide I IR spectrum of a protein also provides information about the backbone conformation. The signal arises mainly from the C=O stretch mode and the N-H bond wagging and bending. Different secondary structures have different peak positions between 1620 and 1690  $\text{cm}^{-1}$ .<sup>27</sup> The amide I band was calculated with Coupled Oscillator Model Spectrum Simulator (COSMOSS).<sup>28</sup> All settings were used at their default values. COSMOSS constructs a vibrational exciton Hamiltonian. The coupling between the local amide I modes was modelled using transition dipole coupling with the nearest neighbour coupling corrected by Jansen's ( $\phi$ ,  $\psi$ ) angle map.<sup>29</sup> The local mode frequencies of the amide I vibrations are also calculated using a nearest-neighbour frequency shift.<sup>29</sup> The calculation of the IR amide I band was based on a single structure, but the influence of conformational dynamics on the inhomogeneous broadening of the bands is approximated by adding some random disorder to the elements of the exciton Hamiltonian: 20  $\text{cm}^{-1}$  to the diagonal terms and 5  $\text{cm}^{-1}$  to the off-diagonal elements, corresponding to the magnitude of fluctuations that might be anticipated in solution under ambient conditions. Homogeneous broadening is approximated by convolving the computed line spectra with a 10  $\text{cm}^{-1}$  linewidth.

Near-UV CD spectra of proteins, on the other hand, contain information regarding the orientations of the aromatic side chains as well as their interactions with the surrounding environment.<sup>18,30,31</sup> Near-UV CD spectra were calculated with the DichroCalc

server,<sup>32</sup> which uses a matrix formulation to represent the exciton coupling of the chromophores in the protein. In the exciton framework an effective Hamiltonian is constructed from a basis of local excitations.<sup>33</sup> The exciton methodology for protein CD calculations has been recently reviewed.<sup>34</sup> The electronic transitions of the aromatic side chain chromophores, phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp), were described via ab initio calculated parameters<sup>35</sup> extended to incorporate the vibrational structure under the electronic bands.<sup>36</sup> Backbone chromophores were modelled with an ab initio parameter set<sup>37</sup> and two transitions were employed for the peptide bond. The calculated near-UV CD line spectra were convolved with a Gaussian function with a 4 nm bandwidth. Calculated intensities from each of the MD snapshots were averaged with equal weighting to give the final spectrum.

## Results

Figure 1 compares the spectroscopic features of the four structures of the main protease in its apo form. All four structures have almost identical far-UV CD spectra with one moderate positive peak at 190 nm, one moderate negative band at 208 nm and an unresolved shoulder at 220 nm.



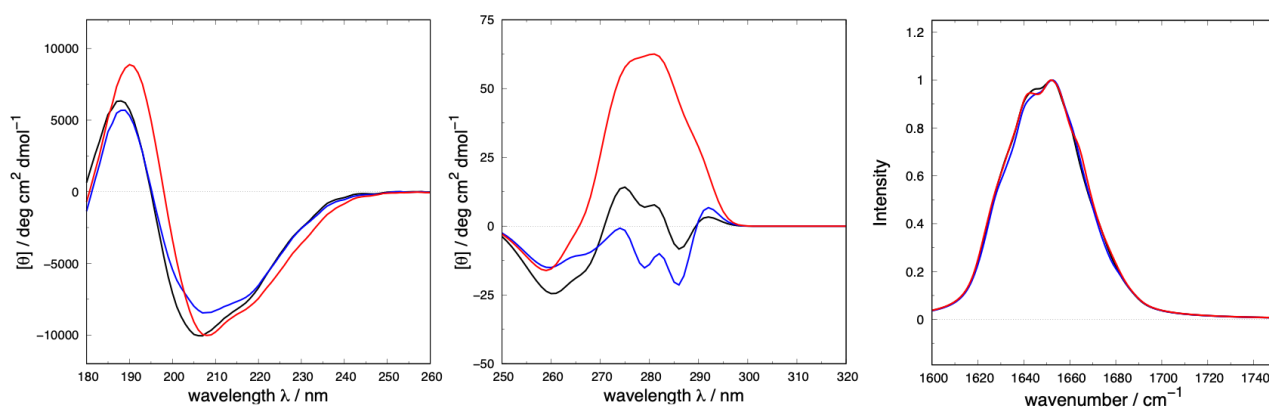
**Figure 1.** Calculated spectra with four experimentally determined structures of the apo-form SARS-CoV-2 main protease. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6M03 (black), 6Y2E (blue), 6Y84 (red), 5R8T (green). The spectrum depicted with dashed lines in the middle panel is calculated with MD simulation snapshots.

This feature matches the mixed  $\alpha$ -helical and  $\beta$ -strand content of the secondary structure of this protein (Table S1). The calculated amide I IR bands of these four structures are very similar as well. The near-UV CD spectra, however, show obvious differences. Calculation with the 6Y2E structure leads to less intense signals in the tyrosine and tryptophan region (270 – 290 nm), while the other three structures show similar bands in terms of the band shape and the intensity. 6Y2E and 6M03 have a stronger negative bands in the phenylalanine region (250 – 268 nm) than the other two 3C-like proteinase structures. This example illustrates that crystal structures from different experiments show similar secondary structure content, but the conformations of the aromatic side chains may be different. Near-UV CD spectrum is sensitive to the different orientation of the aromatic side chains. Calculated spectra of the methyltransferase-nsp10 complex (Figure S1), ADP ribose phosphatase (Figure S2) and RNA polymerase-nsp7-nsp8 complex (Figure S6) also illustrate that the near-UV CD is sensitive to the local conformational differences of the aromatic residues. The other three proteins, nsp9 (Figure S4), nucleocapsid phosphoprotein (Figure S5) and the spike protein RBD and ACE2 complex (Figure S7), have similar band shapes for all three types of calculated spectra.

We have calculated the near-UV CD spectra with the snapshots extracted from the trajectories provided by D. E. Shaw Research<sup>12</sup> (Figure 1, middle panel), since theoretical prediction of the spectrum can sometimes be enhanced by using a set of conformations sampled from an equilibrium ensemble.<sup>38</sup> The calculated spectra with snapshots show similar band shapes for the tyrosine and phenylalanine region with varying magnitude. However, the tryptophan peak is oppositely signed.

Small molecule inhibitors are one possible solution to prevent the function of SARS-CoV-2 main protease in processing RNA translated proteins. Figure 2 compares the calculated spectra of the main protease with and without the inhibitor binding. Compared to the apo form (6Y2E), inhibitor binding structures (6Y2F and 6Y2G) have slightly lower

helical content and a slightly higher amount of turn (Table S1). 6Y2G has the same  $\beta$ -strand content as the apo form, whereas 6Y2E has less  $\beta$ -strand in its structure. These secondary structure changes are reflected in the far-UV CD spectra (Figure 2 left panel). In the near-UV CD spectra (Figure 2 middle panel), binding to inhibitors induced opposite spectroscopic changes. 6Y2F shows a weak band with opposite sign in the tyrosine region (270 – 286 nm) compared to the apo structure (6Y2E), while 6Y2G gives a much stronger positive spectral signal from the tyrosine and tryptophan residues. The influence of inhibitor binding on the near-UV CD arises from the overall conformation changes of the protein rather than direct electronic coupling with the aromatic side chain, since only one phenylalanine is within 6 Å of the inhibitor (Figure S13).

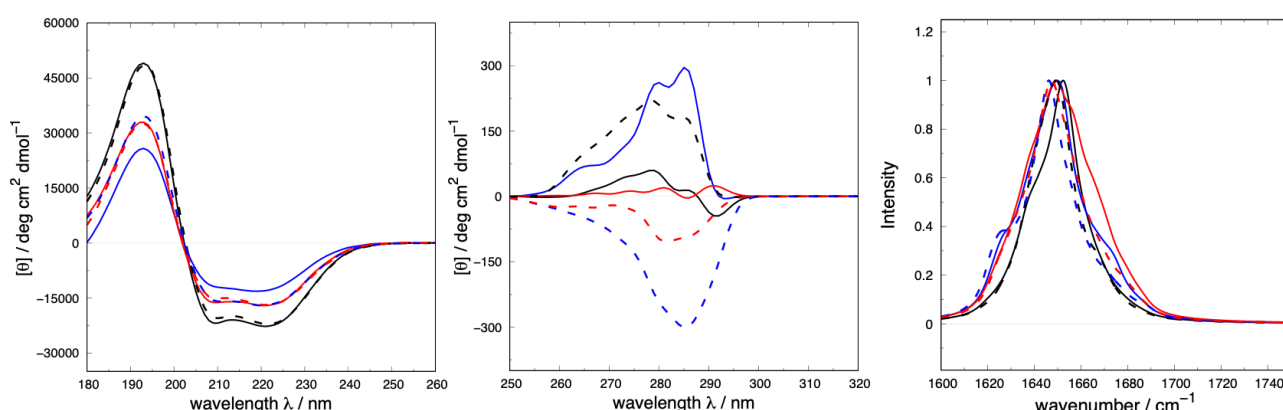


**Figure 2.** Calculated spectra with crystal structures of SARS-CoV-2 main protease with (red and blue) or without (black) alpha-ketoamide inhibitors binding. Left: far-UV CD; middle: near-UV CD; right: IR. PDB ID and colour code: 6Y2E (black), 6Y2F (blue), 6Y2G (red).

Figure 3 shows examples of calculated spectra with different models of the same proteins, nsp6, M-protein and nsp4. The calculated far-UV CD features (Figure 3) consist of an intense positive peak at 190 nm and negative peaks at 208 and 220 nm, which correspond well to this secondary structure composition, namely high helix content (Table S1). The intensities of peaks decrease with the reduction of the helical content of the protein. The two models of nsp6 show a similar band shape in the near-UV CD spectrum below 290 nm, but the Feig model gives weaker calculated signals. The tryptophan in the Feig model leads to a negative signal, while no such band was



observed in the AlphaFold structure. For M-protein and nps4, models from the two groups have opposite signed peaks in their calculated near-UV CD spectra. We are still investigating the origin of this. However, it is hard to dissect the near-UV CD signals especially for proteins with a large number of aromatic residues.



**Figure 3.** Calculated spectra of SARS-CoV-2 proteins with models from the Feig group (solid line) or AlphaFold (dashed line). Left: far-UV CD; middle: near-UV CD; right: IR. Colour code: nsp6 (black), M-protein (blue) and nsp4 (red).

ORF10 and Protein 7a have no helical structure. They contain 26.3% and 60.3%  $\beta$ -strand content, respectively. As shown in Figure S10, they have very different far-UV CD spectra. This is mainly due to the nature of the  $\beta$  structures. Regular  $\beta$  sheets have a positive band at around 195 nm and a negative band with comparable magnitude near 216 nm. A spectrum with a positive band near 190 nm and a minimum at about 200 nm indicates irregular  $\beta$  structures in the protein.<sup>39</sup> Thus, Protein 7a shows regular  $\beta$ -sheet features and ORF10 has more  $\beta$ -bulges and irregular strands in its structure.

## Conclusions

The experimentally determined protein structures share similar secondary structures among different PDB entries. These generated similar calculated far-UV CD and amide I IR spectra. The near-UV CD spectrum is sensitive to the local conformational changes of aromatic residues. Thus, it can distinguish subtly varying tertiary structure features in different protein models. Near-UV CD spectra are difficult to

interpret in the absence of complementary experimental or computational data. However, when combined with an ensemble of MD snapshots, calculated spectra may correspond directly to the actual conformational distribution of the protein in solution. Furthermore, the calculated spectra are sensitive to ligand binding. They provide a direct connection between models and experimental observables and they should be useful in assessing the accuracy of the computational models when compared with experimentally measured spectra of the proteins. The above remarks, of course, apply generally to proteins. In the context of the intense current interest in the SARS-CoV-2 proteins, we suggest that measurement of the optical spectra, particularly of the near-UV CD spectra, would be a valuable complement to the ongoing associated structural, simulation and modelling studies.

#### **Supporting Information Available:**

Table of experimental structures and computational models used. Calculated spectra of all SARF-CoV-2 proteins. Figure showing aromatic residues near the binding pocket of the main protease. The numerical data have been deposited on Mendeley.<sup>40</sup>

#### **References**

- (1) Walls, A. C.; Park, Y-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, **2020**, *180*, 281-292.
- (2) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science*, **2020**, *368*, 409-412.
- (3) Wang, J. Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. *J. Chem. Inf. Model.*, **2020**, *60*, 3277-3286.

- (4) Huynh, T.; Wang, H.; Luan, B. In silico exploration of the molecular mechanism of clinically oriented drugs for possibly inhibiting SARS-CoV-2's main protease. *J. Phys. Chem. Lett.*, **2020**, *11*, 4413-4420.
- (5) Han, Y.; Král, P. Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ACS Nano*, **2020**, *14*, 5143-5147.
- (6) Spinello, A.; Saltalamacchia, A.; Magistrato, A. Is the rigidity of SARS-CoV-2 spike receptor-binding motif the hallmark for its enhanced infectivity? Insights from all-atom simulations. *J. Phys. Chem. Lett.*, **2020**, *11*, 4785-4790.
- (7) Wang, H.; He, S.; Deng, W.; Zhang, Y.; Li, G.; Sun, J.; Zhao, W.; Guo, Y.; Yin, Z.; Li, D.; Shang, L. Comprehensive insights into the catalytic mechanism of Middle East Respiratory Syndrome 3C-like protease and Severe Acute Respiratory Syndrome 3C-like protease. *ACS Catal.* **2020**, *10*, 5871-5890.
- (8) Amaro, R.E.; Mulholland, A.J. A community letter regarding sharing biomolecular simulation data for COVID-19. *J. Chem. Inf. Model.*, **2020**, *60*, 2653-2656.
- (9) <https://swissmodel.expasy.org/repository/species/2697049>.
- (10) Jumper, J.; Tunyasuvunakool, K.; Kohli, P.; Hassabis, D. and the AlphaFold team. DeepMind website, March 5, 2020. <https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19>.
- (11) Heo, L.; Feig, M. bioRxiv 2020, <https://doi.org/10.1101/2020.03.25.008904>.
- (12) D. E. Shaw Research, "Molecular Dynamics Simulations Related to SARS-CoV-2," D. E. Shaw Research Technical Data, 2020. [http://www.deshawresearch.com/resources\\_sarscov2.html](http://www.deshawresearch.com/resources_sarscov2.html).
- (13) Benkert, P.; Biasini, M.; Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, **2011**, *27*, 343-350.
- (14) Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U. S. A.*, **2020**, *117*, 1496-1503.

- (15) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **1983**, 22, 2577-2637.
- (16) Touw, W.G.; Baakman, C.; Black, J.; te Beek, T.A.H.; Krieger, E.; Joosten, R.P.; Vriend, G. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.*, **2015**, 43, D364-D368.
- (17) Chen, Y.H.; Yang, J.T.; Martinez, H.M. Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry*, **1972**, 11, 4120-4131.
- (18) Woody, R.W. Circular dichroism. *Methods Enzymol.*, **1995**, 246, 34-71.
- (19) Lin, M.-H.; Chuang, S.-J.; Chen, C.-C.; Cheng, S.-C.; Cheng, K.-W.; Lin, C.-H.; Sun, C.Y.; Chou, C.-Y. Structural and functional characterization of MERS coronavirus papain-like protease. *J. Biomed. Sci.*, **2014**, 21, 54.
- (20) Li, Y.; Surya, W.; Claudine, S.; Torres, J. Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.*, **2014**, 289, 12535-12549.
- (21) Mycroft-West, C.; Su, D.; Elli, S.; Guimond, S.; Miller, G.; Turnbull, J.; Yates, E.; Guerrini, M.; Fernig, D.; Lima, M.; Skidmore, M. The 2019 coronavirus (SARS-CoV-2) surface protein (Spike) S1 Receptor Binding Domain undergoes conformational change upon heparin binding. bioRxiv 2020, <https://doi.org/10.1101/2020.02.29.971093>.
- (22) Minakshi, R.; Padhan, K.; Rehman, S.; Hassan, I.M.; Ahmad, F. The SARS Coronavirus 3a protein binds calcium in its cytoplasmic domain. *Virus Res.*, **2014**, 191, 180-183.
- (23) Liu, S.; Xiao, G.; Chen, Y.; He, Y.; Niu, J.; Escalante, C.R.; Xiong, H.; Farmer, J.; Debnath, A.K.; Tien, P.; Jiang, S. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet*, **2004**, 363, 938-947.

- (24) Tripet, B.; Howard, M.W.; Jobling, M.; Holmes, R.K.; Holmes, K.V.; Hodges, R.S. Structural characterization of the SARS-coronavirus spike S fusion protein core. *J. Biol. Chem.*, **2004**, 279, 20836-20849.
- (25) Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S. et al., Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.*, **2020**, 30, 343-355.
- (26) Mavridis, L.; Janes, R.W. PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics*, **2017**, 33, 56-63.
- (27) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy *Nat. Protoc.* **2015**, 10, 382-396.
- (28) Ho, J.-J. COSMOSS: Coupled OScillator MOdel Spectrum Simulator.  
<https://github.com/JJ-Ho/COSMOSS> Retrieved April 28, 2020.
- (29) Jansen, T.L.C.; Dijkstra, A.G.; Watson, T.M.; Hirst, J.D. Knoester, J. Modeling the amide I bands of small peptides. *J. Chem. Phys.*, **2006**, 125, 44312.
- (30) Sreerama, N.; Manning, M. C.; Powers, M. E.; Zhang, J. X.; Goldenberg, D. P.; Woody, R. W. Tyrosine, phenylalanine, and disulfide contributions to the circular dichroism of proteins: Circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. *Biochemistry*, **1999**, 38, 10814-10822.
- (31) Karabancheva, T.; Christov, C. Comparative theoretical study of the mechanisms of generation of rotational strengths in the near-UV in  $\beta$ -lactamases from class A. *Chem. Phys. Lett.*, **2004**, 398, 511-516.
- (32) Bulheller, B.M.; Hirst, J.D. DichroCalc–circular and linear dichroism online. *Bioinformatics*, **2009**, 25, 539-540.
- (33) Woody, R. W. Improved calculation of the  $nn^*$  rotational strength in polypeptides. *J. Chem. Phys.* **1968**, 49, 4797–4806.

- (34) Rogers, D. M.; Jasim, S. B.; Dyer, N. T.; Auvray, F.; Rèfrègiers, M.; Hirst, J.D., Electronic circular dichroism of proteins. *Chem*, **5**, 2751–2774 (2019).
- (35) Rogers, D.M.; Hirst, J.D. First principles calculations of protein circular dichroism in the near-ultraviolet. *Biochemistry*, **2004**, *43*, 11092-11102.
- (36) Li, Z.; Hirst, J.D. Vibrational structure in the near-ultraviolet electronic circular dichroism spectra of proteins. *Chem. Sci.*, **2017**, *8*, 4318-4333.
- (37) Besley, N.A.; Hirst, J.D. Theoretical studies toward quantitative protein circular dichroism calculations. *J. Am. Chem. Soc.*, **1999**, *121*, 9636-9644.
- (38) Hirst, J.D.; Bhattacharjee, S.; Onufriev, A.V. Theoretical studies of time-resolved protein folding. *Faraday Discuss.*, **2003**, *122*, 253-267.
- (39) Manavalan, P.; Johnson Jr., W.C. Sensitivity of circular dichroism to protein tertiary structure class. *Nature*, **1983**, *305*, 831-832.
- (40) Hirst, Jonathan (2020), "Computed optical spectra of SARS-CoV-2 proteins", Mendeley Data, V1, doi: 10.17632/zzc4hxxv6p9.1.